Thesis Summary:
Toward a theory of Steganography

Nicholas Hopper

July 14, 2004

Abstract

Informally, *steganography* refers to the practice of hiding secret messages in communications over a public channel so that an eavesdropper (who listens to all communications) cannot even tell that a secret message is being sent. In contrast to the active literature proposing new concrete steganographic protocols and analysing flaws in existing protocols, there has been very little work on formalizing steganographic notions of security, and none giving complete, rigorous proofs of security in a satisfying model.

This thesis initiates the study of steganography from a cryptographic point of view. We give a precise model of a communication channel and a rigorous definition of steganographic security, and prove that relative to a channel oracle, secure steganography exists if and only if one-way functions exist. We give tightly matching upper and lower bounds on the maximum *rate* of any secure stegosystem. We introduce the concept of steganographic key exchange and public-key steganography, and show that provably secure protocols for these objectives exist under a variety of standard number-theoretic assumptions. We consider several notions of *active attacks* against steganography, show how to achieve each under standard assumptions, and consider the relationships between these notions. Finally, we extend the concept of steganograpy as covert communication to include the more general concept of covert *computation*.

1 Introduction

This dissertation focuses on the problem of steganography: how can two communicating entities send secret messages over a public channel so that a third party cannot detect the presence of the secret messages? Notice how the goal of steganography is different from classical encryption, which seeks to conceal the *content* of secret messages: steganography is about hiding the very existence of the secret messages.

Steganographic "protocols" have a long and intriguing history that goes back to antiquity. There are stories of secret messages written in invisible ink or hidden in love letters (the first character of each sentence can be used to spell a secret, for instance). More recently, steganography was used by prisoners, spies and soldiers during World War II because mail was carefully inspected by both the Allied and Axis governments at the time [37]. Postal censors crossed out anything that looked like sensitive information (e.g. long strings of digits), and they prosecuted individuals whose mail seemed suspicious. In many cases, censors even randomly deleted innocent-looking sentences or entire paragraphs in order to prevent secret messages from being delivered. More recently there has been a great deal of interest in digital steganography, that is, in hiding secret messages in communications between computers.

The recent interest in digital steganography is fueled by the increased amount of communication which is mediated by computers and by the numerous potential commercial applications: hidden information could potentially be used to detect or limit the unauthorized propagation of the innocent-looking "carrier" data. Because of this, there have been numerous proposals for protocols to hide data in channels containing pictures [36, 39], video [39, 41, 58], audio [31, 47], and even typeset text [12]. Many of these protocols are extremely clever and rely heavily on domain-specific properties of these channels. On the other hand, the literature on steganography also contains many clever attacks which detect the use of such protocols. In addition, there is no clear consensus in the literature about what it should mean for a stegosystem to be secure; this ambiguity makes it unclear whether it is even possible to have a secure protocol for steganography.

The main goal of this thesis is to rigorously investigate the open question: "under what conditions do secure protocols for steganography exist?" We will give rigorous cryptographic definitions of steganographic security in multiple settings against several different types of adversary, and we will demonstrate necessary and sufficient conditions for security in each setting, by exhibiting protocols which are secure under these conditions.

2 Cryptography and Provable Security

The rigorous study of *provably secure* cryptography was initiated by Shannon [55], who introduced an information-theoretic definition of security: a cryptosystem is secure if an adversary who sees the *ciphertext* - the scrambled message sent by a cryptosystem - receives no additional information about the *plaintext* - the unscrambled content. Unfortunately, Shannon also proved that any cryptosystem which is perfectly secure requires that if a sender wishes to transmit N bits of plaintext data, the sender and the receiver must share at least N bits of random, secret data - the *key*. This limitation means that only parties who already possess secure channels (for the exchange of secret keys) can have secure communications.

To address these limitations, researchers introduced a theory of security against *computationally limited* adversaries: a cryptosystem is computationally secure if an adversary who sees the ciphertext cannot compute (in, e.g. polynomial time) any additional information about the plaintext than he could without the ciphertext[30]. Potentially, a cryptosystem which could be proven secure in this way would allow two parties who initially share a very small number of secret bits (in the case of public-key cryptography, zero) to subsequently transmit an essentially unbounded number of message bits securely.

Proving that a system is secure in the computational sense has unfortunately proved to be an enormous challenge: doing so would resolve, in the negative, the open question of whether $P = NP$. Thus the cryptographic theory community has borrowed a tool from complexity theory: reductions. To prove a cryptosystem secure, one starts with a computational problem which is presumed to be intractible, and a model of how an adversary may attack a cryptosystem, and proves via reduction that computing any additional information from a ciphertext is equivalent to solving the computational problem. Since the computational problem is assumed to be intractible, a computationally limited adversary capable of breaking the cryptosystem would be a contradiction and thus should not exist. In general, computationally secure cryptosystems have been shown to exist if and only if "one-way functions," which are easy to compute but computationally hard to invert, exist. Furthermore, it has been shown that the difficulty of a wide number of well-investigated number-theoretic problems would imply the existence of one-way functions, for example the problem of computing the factors of a product of two large primes [13], or computing discrete logarithms in a finite field[14].

Subsequent to these breakthrough ideas [30, 13], cryptographers have investigated a wide variety of different ways in which an adversary may attack a cryptosystem. For example, he may be allowed to make up a plaintext message and ask to see its corresponding ciphertext, (called a chosen-plaintext attack), or even to make up a ciphertext and ask to see what the corresponding plaintext is (called a chosen-ciphertext attack [46, 49]). Or the adversary may

have a different goal entirely [22, 8, 38] - for example, to modify a ciphertext so that if it previously said "Attack" it now reads as "Retreat" and vice-versa. We will draw on this practice to consider the security of a steganographic protocol under several different kinds of attack.

3 Previous work on theory of steganography

The scientific study of steganography in the open literature began in 1983 when Simmons [56] stated the problem in terms of communication in a prison. In his formulation, two inmates, Alice and Bob, are trying to hatch an escape plan. The only way they can communicate with each other is through a public channel, which is carefully monitored by the warden of the prison, Ward. If Ward detects any encrypted messages or codes, he will throw both Alice and Bob into solitary confinement. The problem of steganography is, then: how can Alice and Bob cook up an escape plan by communicating over the public channel in such a way that Ward doesn't suspect anything "unusual" is going on.

Anderson and Petitcolas [6] posed many of the open problems resolved in this thesis. In particular, they pointed out that it was unclear how to prove the security of a steganographic protocol, and gave an example which is similar to the protocol we present in Chapter 3. They also asked whether it would be possible to have steganography without a secret key, which we address in Chapter 4. Finally, they point out that while it is easy to give a loose upper bound on the rate at which hidden bits can be embedded in innocent objects, there was no known lower bound.

Since the paper of Anderson and Petitcolas, several works [16, 42, 54, 62] have addressed information-theoretic definitions of steganography. Cachin's work [16, 17] formulates the problem as that of designing an encoding function so that the relative entropy between *stegotexts*, which encode hidden information, and independent, identically distributed samples from some innocent-looking *covertext* probability distribution, is small. He gives a construction similar to one we describe, but concludes that it is computationally intractible; and another construction which is provably secure but relies critically on the assumption that all orderings of covertexts are equally likely. Cachin also points out several flaws in other published information-theoretic formulations of steganography.

All information-theoretic formulations of steganography are severely limited, however, because it is easy to show that information-theoretically secure steganography implies information-theoretically secure encryption; thus any secure stegosystem with N bits of secret key can encode at most N hidden bits. In addition, techniques such as public-key steganography and robust steganography are information-theoretically impossible.

4 Contributions of the thesis

The primary contribution of this thesis is a rigorous, cryptographic theory of steganography. The results which establish this theory fall under several categories: symmetric-key steganography, public-key steganography, steganography with active adversaries, steganographic rate, and steganographic *computation*. Here we summarize the results in each category.

Symmetric Key Steganography.

Symmetric-key steganography is the most basic setting for steganography: Alice and Bob possess a shared secret key and would like to use it to exchange hidden messages over a public channel so that Ward cannot detect the presence of these messages. Despite the apparent simplicity of this scenario, there has been little work on giving a precise formulation of steganographic security. Our goal is to give such a formal description.

We first give definitions dealing with the correctness and security of symmetric-key steganography, in terms of indistinguishability from a probabilistic *channel* process $\mathcal{C}$ which models communication as a sequence of *documents* drawn from a set D. Then we show that these notions are *feasible* by giving constructions which satisfy them, under the assumption that pseudorandom function families exist. Finally, we explore the *necessary* conditions for the existence of secure symmetric-key steganography. We show that secure stegosystems relative to a channel exist only if one-way functions exist relative to the channel, and that the existence of a secure stegosystem for a channel implies that the channel is efficiently sampleable.

Public-Key Steganography

Symmetric-key steganograph assumes that the sender and receiver share a secret, randomly chosen key. In the case that some exchange of key material was possible before the use of steganography was necessary, this may be a reasonable assumption. In the more general case, two parties may wish to communicate steganographically, without prior agreement on a secret key. We call such communication *public key steganography*. Whereas previous work has shown that symmetric-key steganography is possible – though inefficient – in an information-theoretic model, public steganography is information-theoretically *impossible*. Thus our complexity-theoretic formulation of steganographic secrecy is crucial to the question of public-key stegangoraphy.

We first introduce some required basic primitives from the theory of public-key cryptography, including the nonstandard notion of a public-key cryptosystem that is indistinguishable from random bits. We then give definitions for public-key steganography and show how to

use these primitives to construct a public-key stegosystem. Finally, we introduce the notion of steganographic key exchange, in which two parties have an innocent looking conversation and at the end, can agree on a key that looks random to any external observer, and give a construction which is secure under the Integer Decisional Diffie-Hellman assumption.

Steganography with active adversaries

The previously described results show that a *passive* adversary (one who simply eavesdrops on the communications between Alice and Bob) cannot hope to subvert the operation of a stegosystem. In this chapter, we consider the notion of an *active* adversary who is allowed to introduce new messages into the communications channel between Alice and Bob. In such a situation, an adversary could have two different goals: disruption or detection.

Disrupting adversaries attempt to prevent Alice and Bob from communicating steganographically, subject to some set of publicly-known restrictions. We call a stegosystem which is secure against this type of attack *robust*. We will give a formal definition of *robustness* against such an attack, consider what type of restrictions on an adversary are *necessary* for the existence of a robust stegosystem, and give the first construction of a provably robust stegosystem against any set of restrictions satisfying this necessary condition. Our protocol is secure assuming the existence of pseudorandom functions.

Distinguishing adversaries introduce additional traffic between Alice and Bob in hopes of tricking them into revealing their use of steganography. We consider the security of symmetric- and public-key stegosystems against active distinguishers, and give constructions which are secure against various notions of active distinguishing attacks. In order to do so, we introduce the notion of a cryptoystem which is indistinguishable from random bits under adaptive chosen ciphertext attack, and exhibit symmetric-key and public-key cryptosystems satisfying this notion.

We also show that *no stegosystem can be simultaneously secure against both disrupting and distinguishing active adversaries*. This contradicts a conjecture that the two goals can be addressed orthogonally, stated in a recent paper [7] which addresses the issue of active distinguishing adversaries.

Bounds on steganographic rate

Intuitively, the *rate* of a stegosystem is the number of bits of hiddentext that a stegosystem encodes per document of covertext. Clearly, for practical use a stegosystem should have a relatively high rate, since it may be impractical to send many documents to encode just a few bits. Thus an important question for steganography, first posed by Anderson and Petitcolas [6] is "how much information can be safely encoded by a stegosystem in the channel $\mathcal{C}$?"

A trivial upper bound on the rate of a stegosystem is $\log |D|$. Prior to our work, there were no provably secure stegosystems, and so there was no known lower bound. The rate of our previous constructions is $o(1)$, that is, as the security parameter k goes to infinity, the rate goes to 0. In this chapter, we will address the question of what the optimal rate is for a (universal) stegosystem. We first formalize the definition of the rate of a universal stegosystem. We will then tighten the trivial upper bound by giving a rate MAX such that any universal stegosystem with rate exceeding MAX is insecure. We will then give a matching lower bound by exhibiting a provably secure stegosystem with rate $(1-o(1))MAX$. Finally we will address the question of what rate a robust stegosystem may achieve: we give an upper bound $RMAX$ above which a universally robust stegosystem is insecure, and a construction with rate $(1-\epsilon)RMAX$ for any $\epsilon > 0$.

Covert Computation

We introduce the novel concept of *covert two-party computation.* Whereas ordinary secure two-party computation only guarantees that no more knowledge is leaked about the inputs of the individual parties than the result of the computation, covert two-party computation employs steganography to yield the following additional guarantees: (A) no outside eavesdropper can determine whether the two parties are performing the computation or simply communicating as they normally do; (B) before learning $f(x_A, x_B)$, neither party can tell whether the other is running the protocol; (C) after the protocol concludes, each party can only determine if the other ran the protocol insofar as they can distinguish $f(x_A, x_B)$ from uniformly chosen random bits. Covert two-party computation thus allows the construction of protocols that return $f(x_A, x_B)$ only when it equals a certain value of interest (such as "Yes, we are romantically interested in each other") but for which *neither party can determine whether the other even ran the protocol whenever $f(x_A, x_B)$ does not equal the value of interest.* We introduce security definitions for covert two-party computation and we construct protocols with provable security based on the Decisional Diffie-Hellman assumption.

References

[1] G. Aggarwal, N. Mishra and B. Pinkas. Secure computation of the k'th-ranked element To appear in *Advances in Cryptology – Proceedings of Eurocrypt '04*, 2004.

[2] Luis von Ahn, Manuel Blum and John Langford. Telling Humans and Computers Apart (Automatically) or How Lazy Cryptographers do AI.

[3] Luis von Ahn and Nicholas J. Hopper. Public-Key Steganography. Submitted to CRYPTO 2003.

[4] L. von Ahn and N. Hopper. Public-Key Steganography. To appear in *Advances in Cryptology – Proceedings of Eurocrypt '04*, 2004.

[5] Ross J. Anderson and Fabien A. P. Petitcolas. *On The Limits of Steganography.* IEEE Journal of Selected Areas in Communications, 16(4). May 1998.

[6] Ross J. Anderson and Fabien A. P. Petitcolas. *Stretching the Limits of Steganography.* In: *Proceedings of the first International Information Hiding Workshop.* 1996.

[7] M. Backes and C. Cachin. Public-Key Steganography with Active Attacks. *IACR e-print archive report 2003/231*, 2003.

[8] M. Bellare, A. Desai, D. Pointcheval, and P. Rogaway. Relations Among Notions of Security for Public-Key Encryption Schemes. In: *Advances in Cryptology – Proceedings of CRYPTO 98*, pages 26–45, 1998.

[9] M. Bellare and P. Rogaway. Random Oracles are Practical. *Computer and Communications Security: Proceedings of ACM CCS 93*, pages 62–73, 1993.

[10] M. Bellare and S. Micali. Non-interactive oblivious transfer and applications. *Advances in Cryptology – Proceedings of CRYPTO '89*, pages 547-557, 1990.

[11] E.R Berlekamp. Bounded Distance +1 Soft-Decision Reed-Solomon Decoding. *IEEE Transactions on Information Theory*, 42(3), pages 704–720, 1996.

[12] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman. Hiding Information in Documents Images. In: Conference on Information Sciences and Systems, 1995.

[13] M. Blum and S. Goldwasser. An Efficient Probabilistic Public-Key Encryption Scheme Which Hides All Partial Information. *Advances in Cryptology: CRYPTO 84*, Springer LNCS 196, pages 289-302. 1985.

[14] M. Blum and S. Micali. How to generate cryptographically strong sequences of random bits. In: *Proceedings of the 21st FOCS*, pages 112–117, 1982.

[15] E. Brickell, D. Chaum, I. Damgärd, J. van de Graaf: Gradual and Verifiable Release of a Secret. *Advances in Cryptology – Proceedings of CRYPTO '87*, pages 156-166, 1987.

[16] C. Cachin. *An Information-Theoretic Model for Steganography.* In: *Information Hiding – Second International Workshop, Preproceedings.* April 1998.

[17] C. Cachin. *An Information-Theoretic Model for Steganography.* In: *Information and Computation* 192 (1): pages 41–56, July 2004.

[18] R. Canetti, U. Feige, O. Goldreich and M. Naor. Adaptively Secure Multi-party Computation. *28th Symposium on Theory of Computing (STOC 96)*, pages 639-648. 1996.

[19] R. Cramer and V. Shoup. A practical public-key cryptosystem provably secure against adaptive chosen ciphertext attack. *Advances in Cryptology: CRYPTO 98*, Springer LNCS 1462, pages 13-27, 1998.

[20] R. Cramer and V. Shoup. Universal Hash Proofs and a Paradigm for Adaptive Chosen Ciphertext Secure Public-Key Encryption. *Advances in Cryptology: EUROCRYPT 2002*, Springer LNCS 2332, pages 45-64. 2002.

[21] S. Craver. *On Public-Key Steganography in the Presence of an Active Warden.* In: *Information Hiding – Second International Workshop, Preproceedings.* April 1998.

[22] D. Dolev, C. Dwork, and M. Naor. Non-malleable Cryptography. *23rd Symposium on Theory of Computing (STOC '91)*, pages 542-552. 1991.

[23] Z. Galil, S. Haber, M. Yung. Cryptographic Computation: Secure Fault-Tolerant Protocols and the Public-Key Model. *Advances in Cryptology – Proceedings of CRYPTO '87*, pages 135-155, 1987.

[24] O. Goldreich. *Foundations of Cryptography: Basic Tools.* Cambridge University Press, 2001.

[25] O. Goldreich. Secure Multi-Party Computation. Unpublished Manuscript. *http://philby.ucsd.edu/books.html*, 1998.

[26] O. Goldreich, S. Goldwasser and S. Micali. How to construct pseudorandom functions. *Journal of the ACM*, vol 33, 1998.

[27] O. Goldreich and L.A. Levin. A Hardcore predicate for all one-way functions. In: *Proceedings of 21st STOC*, pages 25–32, 1989.

[28] O. Goldreich, S. Micali and A. Wigderson. How to Play any Mental Game. *Nineteenth Annual ACM Symposium on Theory of Computing*, pages 218-229.

[29] S. Goldwasser and M. Bellare. Lecture Notes on Cryptography. Unpublished manuscript, August 2001. available electronically at `http://www-cse.ucsd.edu/~mihir/papers/gb.html`.

[30] S. Goldwasser and S. Micali. Probabilistic Encryption & how to play mental poker keeping secret all partial information. In: *Proceedings of the 14th STOC*, pages 365–377, 1982.

[31] D. Gruhl, W. Bender, and A. Lu. Echo Hiding. In: *Information Hiding: First International Workshop*, pages 295–315, 1996.

[32] J. Hastad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4), pages 1364-1396, 1999.

[33] N. Hopper, J. Langford and L. Von Ahn. Provably Secure Steganography. *Advances in Cryptology – Proceedings of CRYPTO '02*, pages 77-92, 2002.

[34] Nicholas J. Hopper, John Langford, and Luis von Ahn. *Provably Secure Steganography.* CMU Tech Report CMU-CS-TR-02-149, 2002.

[35] Russell Impagliazzo and Michael Luby. *One-way Functions are Essential for Complexity Based Cryptography.* In: 30th FOCS, November 1989.

[36] G. Jagpal. *Steganography in Digital Images* Thesis, Cambridge University Computer Laboratory, May 1995.

[37] D. Kahn. *The Code Breakers.* Macmillan 1967.

[38] J. Katz and M. Yung. Complete characterization of security notions for probabilistic private-key encryption. In: *Proceedings of 32nd STOC*, pages 245–254, 1999.

[39] Stefan Katzenbeisser and Fabien A. P. Petitcolas. *Information hiding techniques for steganography and digital watermarking.* Artech House Books, 1999.

[40] Y. Lindell. A Simpler Construction of CCA2-Secure Public Key Encryption. *Advances in Cryptology: EUROCRYPT 2003*, Springer LNCS 2656, pages 241-254. 2003.

[41] K. Matsui and K. Tanaka. *Video-steganography.* In: *IMA Intellectual Property Project Proceedings*, volume 1, pages 187–206, 1994.

[42] T. Mittelholzer. *An Information-Theoretic Approach to Steganography and Watermarking* In: *Information Hiding – Third International Workshop.* 2000.

[43] M. Naor and B. Pinkas. Efficient Oblivious Transfer Protocols. In: *Proceedings of the 12th Annual ACM/SIAM Symposium on Discrete Algorithms (SODA 2001)*, pages 448–457. 2001.

[44] M. Naor, B. Pinkas and R. Sumner. Privacy Preserving Auctions and Mechanism Design. In: *Proceedings, 1999 ACM Conference on Electronic Commerce.*

[45] M. Naor and M. Yung. Universal One-Way Hash Functions and their Cryptographic Applications. *21st Symposium on Theory of Computing (STOC 89)*, pages 33-43. 1989.

[46] M. Naor and M. Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. *22nd Symposium on Theory of Computing (STOC 90)*, pages 427-437. 1990.

[47] C. Neubauer, J. Herre, and K. Brandenburg. Continuous Steganographic Data Transmission Using Uncompressed Audio. In: *Information Hiding: Second International Workshop*, pages 208–217, 1998.

[48] B. Pinkas. Fair Secure Two-Party Computation. In: *Advances in Cryptology – Eurocrypt '03*, pp 87–105, 2003.

[49] C. Rackoff and D. Simon. Non-interactive Zero-Knowledge Proof of Knowledge and Chosen Ciphertext Attack. *Advances in Cryptology: CRYPTO 91*, Springer LNCS 576, pages 433-444, 1992.

[50] L. Reyzin and S. Russell. Simple Stateless Steganography. IACR e-print archive report 2003/093, 2003.

[51] Phillip Rogaway, Mihir Bellare, John Black and Ted Krovetz. *OCB: A Block-Cipher Mode of Operation for Efficient Authenticated Encryption.* In: *Proceedings of the Eight ACM Conference on Computer and Communications Security (CCS-8).* November 2001.

[52] J. Rompel. One-way functions are necessary and sufficient for secure signatures. *22nd Symposium on Theory of Computing (STOC 90)*, pages 387-394. 1990.

[53] A. Sahai. Non-Malleable Non-Interactive Zero Knowledge and Adaptive Chosen-Ciphertext Security. *40th IEEE Symposium on Foundations of Computer Science (FOCS 99)*, pages 543-553. 1999.

[54] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger *Information theoretic analysis of Steganography.* In: *Proceedings ISIT '98.* 1998.

[55] C.E. Shannon. *Communication theory of secrecy systems.* In: *Bell System Technical Journal*, 28 (1949), pages 656-715.

[56] G.J. Simmons. *The Prisoner's Problem and the Subliminal Channel.* In: *Proceedings of CRYPTO '83.* 1984.

[57] L. Welch and E.R. Berlekamp. Error correction of algebraic block codes. US Patent Number 4,663,470, December 1986.

[58] A. Westfeld, G. Wolf. *Steganography in a Video Conferencing System*. In: *Information Hiding – Second International Workshop, Preproceedings*. April 1998.

[59] A. C. Yao. Protocols for Secure Computation. *Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science*, 1982, pages 160–164.

[60] A. C. Yao. How to Generate and Exchange Secrets. *Proceedings of the 27th IEEE Symposium on Foundations of Computer Science*, 1986, pages 162–167.

[61] A. Young and M. Yung. Kleptography: Using Cryptography against Cryptography. *Advances in Cryptology: Eurocrypt 87*, Springer LNCS 1233, pages 62-74, 1987.

[62] J Zollner, H.Federrath, H.Klimant, A.Pftizmann, R. Piotraschke, A.Westfield, G.Wicke, G.Wolf. *Modeling the security of steganographic systems*. In: *Information Hiding – Second International Workshop, Preproceedings.* April 1998.